PCT
WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : G06F 12/00 — A1

(11) International Publication Number: WO 99/50751

(43) International Publication Date: 7 October 1999 (07.10.99)

(21) International Application Number: PCT/US98/20596

(22) International Filing Date: 30 September 1998 (30.09.98)

(30) Priority Data:

9806973.5	1 April 1998 (01.04.98)	GB
9806977.6	1 April 1998 (01.04.98)	GB
9807001.4	1 April 1998 (01.04.98)	GB

(71) Applicant *(for all designated States except US)*: XEROX CORPORATION [US/US]; Xerox Square 020, Rochester, NY 14644 (US).

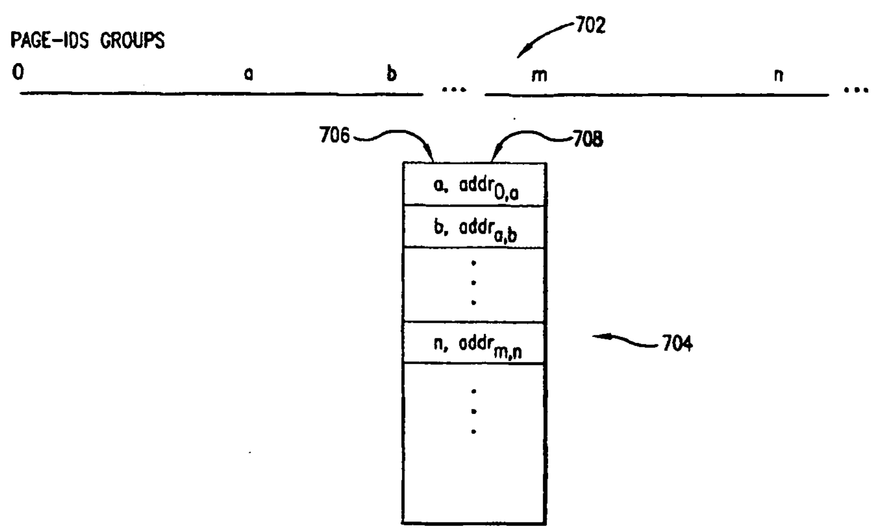
(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: DYMETMAN, Marc [FR/FR]; 3, rue Pierre Duclot, F-38000 Grenoble (FR). COPPERMAN, Max [US/US]; 233 Sunset Avenue, Santa Cruz, CA 95060 (US).

(74) Agents: OLIFF, James, A. et al.; Oliff & Berridge, PLC, P.O. Box 19928, Alexandria, VA 22320 (US).

(81) Designated States: JP, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: ROUTING DOCUMENT IDENTIFIERS



CENTRAL ROUTER'S TABLE

(57) Abstract

Routing is performed by receiving an item of data that indicates a page identifier (702) and by using the page identifier indicated by the item of data to obtain a network address (708).

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania
AM	Armenia
AT	Austria
AU	Australia
AZ	Azerbaijan
BA	Bosnia and Herzegovina
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CU	Cuba
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GH	Ghana
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IL	Israel
IS	Iceland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgyzstan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakstan
LC	Saint Lucia
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
LS	Lesotho
LT	Lithuania
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
MK	The former Yugoslav Republic of Macedonia
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
MX	Mexico
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SG	Singapore
SI	Slovenia
SK	Slovakia
SN	Senegal
SZ	Swaziland
TD	Chad
TG	Togo
TJ	Tajikistan
TM	Turkmenistan
TR	Turkey
TT	Trinidad and Tobago
UA	Ukraine
UG	Uganda
US	United States of America
UZ	Uzbekistan
VN	Viet Nam
YU	Yugoslavia
ZW	Zimbabwe

Routing document identifiers

The present invention relates to functionalities of computers and computer networks, and more particularly to the routing of document identifiers across such networks.

The invention addresses problems arising in the implementation of techniques described in GB patent application 98_______(applicants' ref. R/98003/JDR) based on substrates, e.g. paper, or documents produced therefrom, which include visible or invisible coded markings identifying the substrate and preferably locations or zones within it. This marking scheme in turn preferably uses Xerox DataGlyphs. Such substrates are referred to hereinafter as "coded substrates".

Each physical coded substrate contains a pid-code (pid stands for page-identifier) which identifies it uniquely world-wide and permits to locate the "digital-page" coupled with this physical page, which can sit anywhere on the global network. This pid is encoded in DataGlyphs (visibly or invisibly) on the surface of the page in such a way that a "pointer" equipped with a small camera can recover the pid by looking at a small circular area of radius r, anywhere on the page. With the coded substrates, the space needed for encoding on a physical page the net address of the digital-page counterpart is at a high premium.

Because the pid must be recovered from a small area, it is important to ensure that two conditions are met:

1. The pid is encoded using a small number of bits.

2. The pid can address unambiguously any of a large number of digital-pages.

In order to respect these two conditions, the theoretically optimal scheme is to use a small number of bits for the pid, say 64, and to use the pid to address 2^64 different digital-pages. (To give an idea of how big this figure is: if every inhabitant of the Earth was to produce 80 thousands sheets of Intelligent Paper a day for the next century, a 64 bits pid would be sufficient to uniquely identify all the digital-pages needed.)

Obviously, with such a scheme, there must be a way to map the pid recovered by the pointer into the net address of the corresponding digital page. The solution that is envisaged in the abovementioned patent application (ref.

R/98003/JDR) is to use a central router. This router contains a table of pairs (pid, address). The pointer sends a pid to the router and gets back the address to which it can then connect to retrieve the relevant digital-page.

This centralised routing scheme has two problems, though. First, it may require huge tables for storing the (pid, address) pairs. Secondly, the number of requests per day to the router can be very large.

There is a need for techniques for implementing a workable routing system, without incurring unrealistic address-storage and traffic-frequency costs at the central router's site.

The present invention provides a method carried out in a data processing device comprising receiving a data set, the data set including a page identification code, sending the page identification code to a router, and receiving back from the router a network address, the network address being the address of a server associated with said page identification code.

The invention further provides a method carried out in a data processing system, comprising: receiving a data set from a remote device via network, the data set defining a page identification code, using association data, the association data defining a mapping between a plurality of page identification codes and a plurality of network addresses, determining a network address associated with the received page identification code, and transmitting the network address to the remote device.

The invention further provides a programmable data processing device when suitably programmed for carrying out the methods as described above.

The invention has the advantage of requiring smaller and more efficient routing facilities, and which capitalises on the tendency of publishers (responsible for printing on coded substrates) to buy coded substrates in bulk. The scheme has also the advantage that no address space is lost (64 bits of data on the paper still allow to address 2^64 digital pages).

For the purpose of illustration it is assumed that the company producing (printing) the coded substrates, providing the central address routing and selling the pointers is one and the same company, which we will call company X. It is also assumed that this company sells the sheets, not directly to end users, but to publishers (for instance book or journal publishers) who sell the printed pages (i.e. after printing with human-readable information, such as the text of an article) to the end-users.

(One centralised router is assumed for simplicity. A variant where several mirror sites for the router is obviously compatible with the invention. More specifically: In the non-variant, there is one central router whose address is *a priori* stored in each pointer. Requests are sent to this router by the pointer.

In the variant, there are several (let's say 10 for example) mirror routers that each contain the same information as the single router in the non-variant. They are just copies of this router, but are located at different geographical locations. The addresses of these 10 routers are *a priori* known to the pointers. When the pointer needs to make a request to the router, it is indifferent which router copy it chooses to send this request. The pointer chooses this router randomly among the 10 it knows. If all pointers do the same, this has the net effect of dividing by 10 the number of requests that each router has to answer on average each day. This has also the effect of making a more efficient use of the network communication channels.)

Further advantages of the invention are that it allows a small number of bits encoded on the page to address a large number of digital pages, that it minimises hits on central router, and no redundancy in the address codes is needed (compare hierarchical internet addresses).

It will be appreciated that the techniques described herein may also be used in conjunction with the techniques described in GB patent applications 98______(applicants' ref.R/98003/JDR) and 98______(applicants' ref.R/98004/JDR), filed concurrently herewith.

Embodiments of the invention will now be described, by way of example, with reference to the accompanying drawings, in which:

Figure 1 illustrates the components of a pointed document as printed on a coded substrate;

Figure 2 shows a sample of zones, and the disposition of machine readable data, on a coded substrate;

Figures 3 and 4 show how digital data is encoded in the zones illustrated in Fig. 2;

Figure 5 schematically illustrates an embodiment of a pointer which may be used in implementing the invention;

Figure 6 shows a configuration for passing page identification codes and/or page location codes from the pointer of Fig. 5 to a network computer, in accordance with an embodiment of the invention;

Figure 7 illustrates the assignment of page-id groups and the association of page-ids and server addresses in an embodiment employing centralised routing;

Figure 8 shows the retrieval of the internal address for a first page using a first page-id;

Figure 9 shows the retrieval of the internal address for a second page using a second page-id; and

Figure 10 is a flow chart of the processing steps in implementing the retrieval scheme of Fig. 8.

Figure 1 illustrates the components of a pointed document as printed on a coded substrate. The printed document 102 comprises a layer 104 of printed visible (human-readable) information printed on a coded substrate 106. The coded substrate 106 in turn comprises a layer 108 of visible or invisible machine readable markings printed on a sheet medium 110 (e.g. paper).

Figure 2 shows a sample of zones, and the disposition of machine readable data, on a coded substrate. Each zone or cell 202 includes a border 204 and an orientation marker 206. A first set of markings 208 over part of the interior of the cell 202 are encoded representations of the page-id, while a second set of markings 210 over a (smaller) part of the interior of the cell 202 are encoded representations of the localisation (page-loc) - uniquely defining the position of the cell 202 within the page.

Figures 3 and 4 show how digital data is encoded in the zones illustrated in Fig. 2. Fig. 3 shows the binary data, i.e. 47 bits of page-id in the upper section 302 (the bit stream wraps at the cell border 204), and 16 its of page localisation data (loc) in the lower section 304). The page-id code denotes 108429159095492 = 629DA182DCC4 (hexadecimal) = 11000101001110110100001100000101101110011000100000010000000010101 (binary; to make the wrapping explicit). In the 16 bit loc code in section 304, there are 8 bits for the X co-ordinate and 8 bits for the Y co-ordinate. Thus, for the cell (zone) shown, its position is 16,21 on the substrate.

Fig. 4 shows the same data as in Fig. 3, but represented by Data Glyph markings. Encoding using data glyphs and the retrieval of data therefrom is discussed

further in US-A-5,486,686, EP-A-469864, and the abovementioned GB application (ref.R/98003/JDR). Here, there is a first set of glyphs (markings) in upper section 402 and a second set in lower section 404, the two sets of glyphs being encoded representations of page-id and loc codes.

Figure 5 schematically illustrates an embodiment of a pointer which may be used in implementing the invention. The pointer 502 comprises a marking device 504 (which may be a pen or any other marking device suitable for making marks which are visible to a user), and an image capture device 506. In use, whether or not the user is making marks using the marking device 504, the image capture device 506 is able to capture images of an area A of a document 508. (For the sake of illustration, the sizes of these elements are exaggerated - e.g. in practice, the area A may be much closer to the tip 505 of the marking device 504 than appears). In certain embodiments, the marking device 504 may be omitted.

The document 508 may be a 'blank' coded substrate, or such a substrate having human-readable information printed thereon.

Figure 6 shows a configuration for passing page identification codes and/or page location codes from the pointer of Fig. 5 to a network computer, in accordance with an embodiment of the invention. The image capture device (e.g. CCD camera) 506 is coupled by wired or wireless (e.g. IR or RF) link to processing device 602 and in use provides image data defining capture images to the processing device 602. The operative elements of the processing device 602 are a frame grabber circuit 604, image decoding software 606, and a CPU 608, which are known in the art. (In certain embodiments, the camera 506 and processing device 602 may be combined into an integral handheld unit). In use, the processing device 602 extracts from the image data the corresponding page-id and page-location data (<pid, loc>) and communicates them in a wired or wireless fashion to a local device (here, a network computer 610, which is linked to the network (intranet, internet) in a known manner). The computer 610 has its own unique network address, but need not have any information output device (e.g. display screen, printer).

Figure 7 illustrates the assignment of page-id groups and the association of page-ids and server addresses in an embodiment employing centralised routing. This shows the distribution 702 of page-ids, and the groups thereof (e.g. O-a, a-b), each of which is encoded in one coded substrate of a batch as it is supplied to a publisher for

printing. A table 704 is stored at the central router computer: this stores the associated between groups 706 of consecutive page-ids and server addresses 708. Here $addr_{k,l}$ is the internet address of the server for page-ids k through l.

Figure 8 shows the retrieval of the internal address for a first page using a first page-id.

The pointer 502, while pointed at a coded substrate, can communicate a page-id i to a central router 802 and to a server 804. The protocol may be described as follows (this is illustrated in Fig. 10).

1. Pointer reads page whose page-id is i where $m \leq i < n$.

2. Pointer transmits i to central router.

3. Central router transmits (m, n, $addr_{m,n}$) to pointer.

4. Triple (m, n, $addr_{m,n}$) is stored in pointer.

5. Pointer transmits i to server at $addr_{m,n}$.

6. Server transmits internet address of digital page for i to pointer.

7. Pointer interacts with digital page.

Figure 9 shows the retrieval of the internal address for a second page using a second page-id. Here the pointer 502 communicates page-id j to the server 804. The protocol is as follows.

1. Pointer reads page whose page-id is j where $m \leq j < n$.

2. Pointer transmits j to server at $addr_{m,n}$.

3. Server transmits internet address of digital page for j to pointer.

4. Pointer interacts with digital page.

It can be seen that techniques according to the invention may be used to implement the following scheme.

1. Publisher P buys p coded substrates from X. P provides the net address A of a server belonging to P.

2. X has previously altogether sold m coded substrates to other publishers (or to X). These sheets have been given page-ids (pid's) ranging from 0 to m-1. X produces p new coded substrates, numbered from m to n = m + p - 1.

3. X installs number m as a key in its central router database (it is assumed that number m was previously installed in a similar way), and associates with this key the address addrm,n = A provided by P for these p pages (see Fig. 8).

4. At a later time, a user clicks his/her pointer at one of these p sheets for the first time. This results in the pid of this sheet to be sent to the central router. The central router returns to the pointer a record (m,n,addrm,n), where m, n, and addrm,n are as above. This triple is stored in the pointer's memory (see Fig. 8). (It is essential that pointers be co-operative in storing these triples. This behaviour can be enforced through *a priori* specification or through some "punishing" scheme for uncooperative pointers. Also, because the pointer has limited memory, it may have to expel a previously stored record. This qualification does not seriously affect the proposal.)

5. The pointer sends the pid i'(a number comprised between m and n) to the address addrm,n. The final routing of the pid is now the responsibility of the publisher's server at addrm,n, who retrieves the digital page associated with pid (see Fig. 8).

6. If,' at a later stage, the same pointer clicks on *any* page having a pid j in the range [m,n[, then the pointer consults its memory first, and notices that it contains the record (m,n,addrm,n). Rather than now consulting the central router, it consults directly the publisher's server at address addrm,n (see Fig. 9).

There are two main advantages:

1. In the central router, only one entry needs to be stored for each (batch) bulk purchase of coded substrates.

2. If a journal publisher, let's say, buys in a single purchase enough sheets for a year of publication, a subscriber to the journal will only access the central router once in a year; for the first request to the router will result in caching the whole range of pages for a year of the journal. Because of this tendency for users to click repeatedly on pages belonging to the same sheet batch, the number of per day requests to the central router diminishes dramatically.

The routing fee charged by X to the publishers can promote the scheme by being regressive relative to the number of sheets bought in a single purchase by a publisher. The rationale behind this regressive cost is the fact that the routing costs associated to two separate purchases by the same publisher of p and then p' sheets, associated with two publisher servers A and A', are higher than the routing costs associated with a single purchase of p+p' sheets associated with a single publisher server A. This difference is especially large if the fact that a pointer has ``seen" the first batch significantly increases its probability of seeing the second one sometime in

the future (as compared to the *a priori* probability of seeing the second batch). It will be appreciated that these principles could be repeated recursively: the publisher's server itself could be organised in a way similar to the central router.

CLAIMS:

1. A method for performing routing comprising:

receiving an item of data that indicates a page identifier; and

using the page identifier indicated by the item of data to obtain a network address.

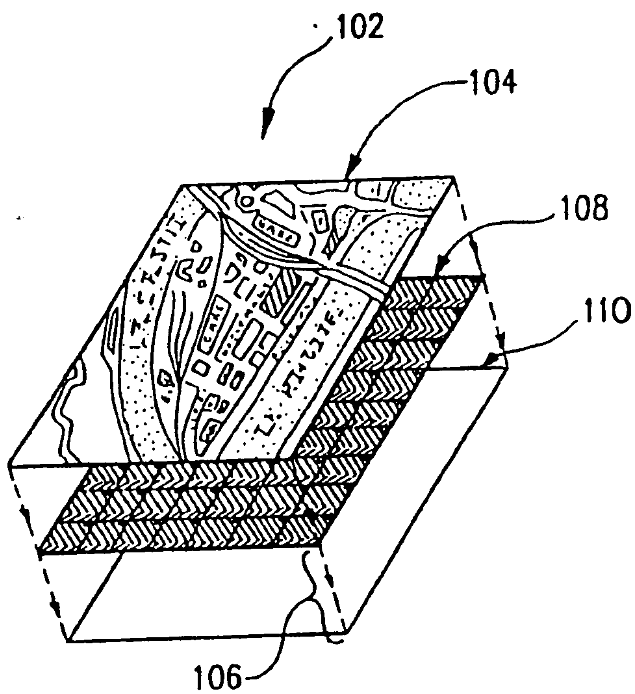


FIG.1

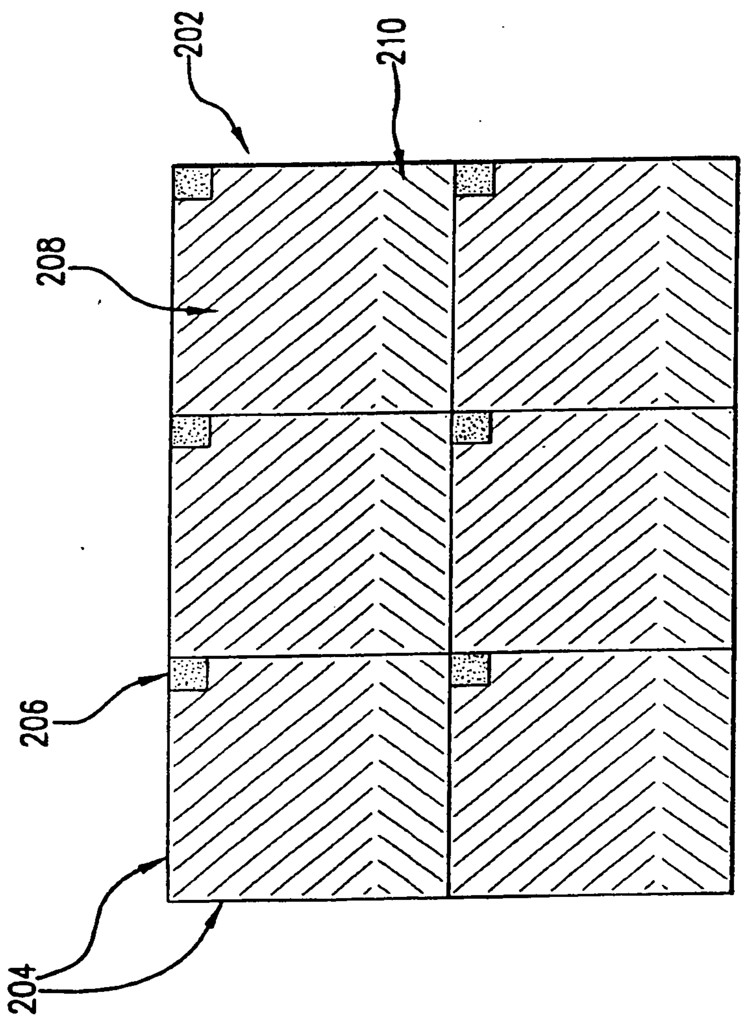
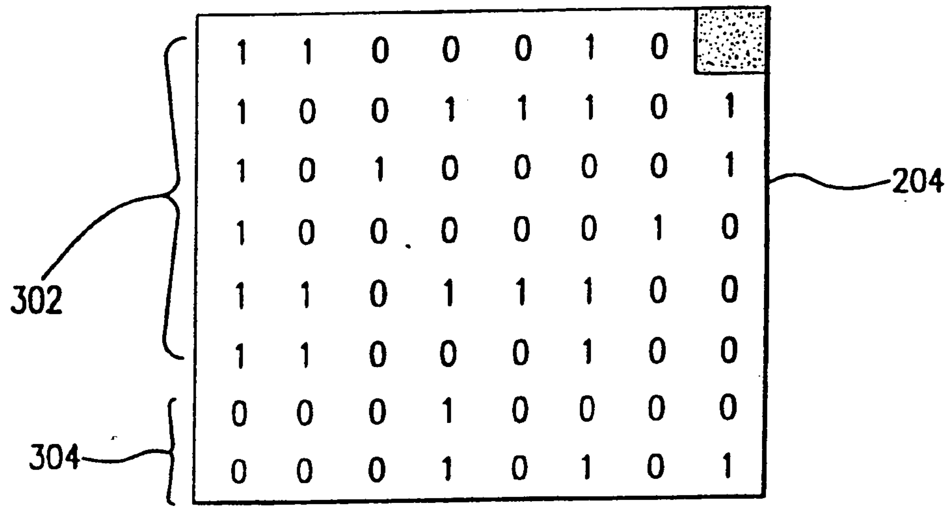
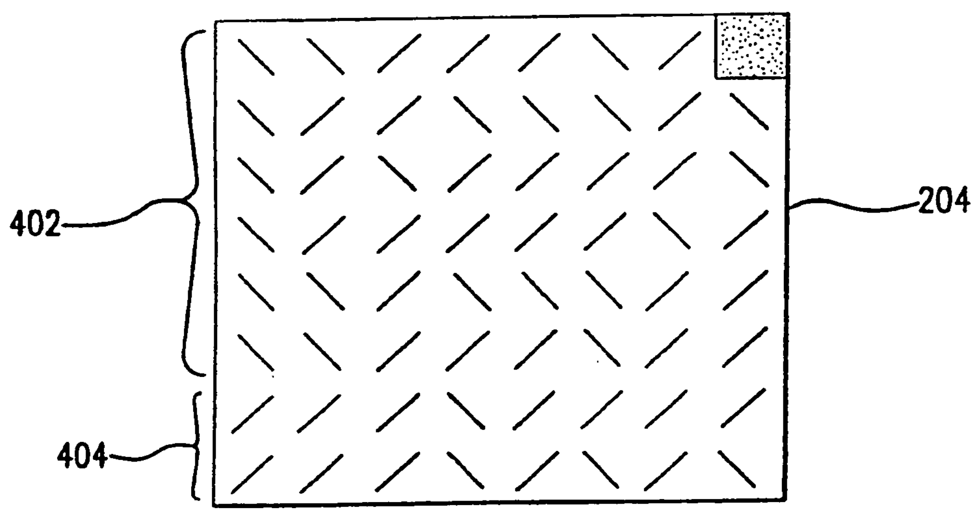


FIG.2



1 1 0 0 0 1 0
1 0 0 1 1 1 0 1
1 0 1 0 0 0 0 1
1 0 0 0 0 0 1 0
1 1 0 1 1 1 0 0
1 1 0 0 0 1 0 0
0 0 0 1 0 0 0 0
0 0 0 1 0 1 0 1
302
304
204

FIG.3



402
404
204

FIG.4

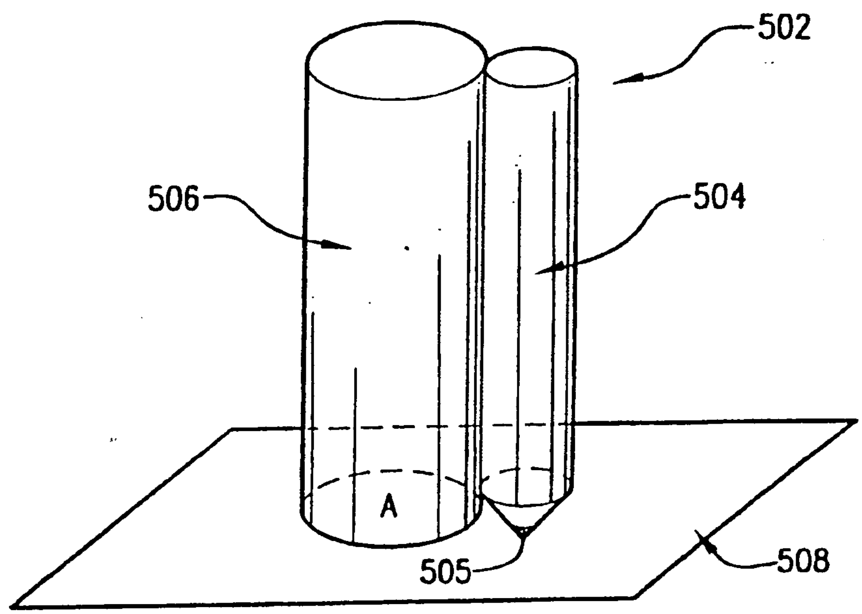

502
506
504
A
505
508

FIG.5

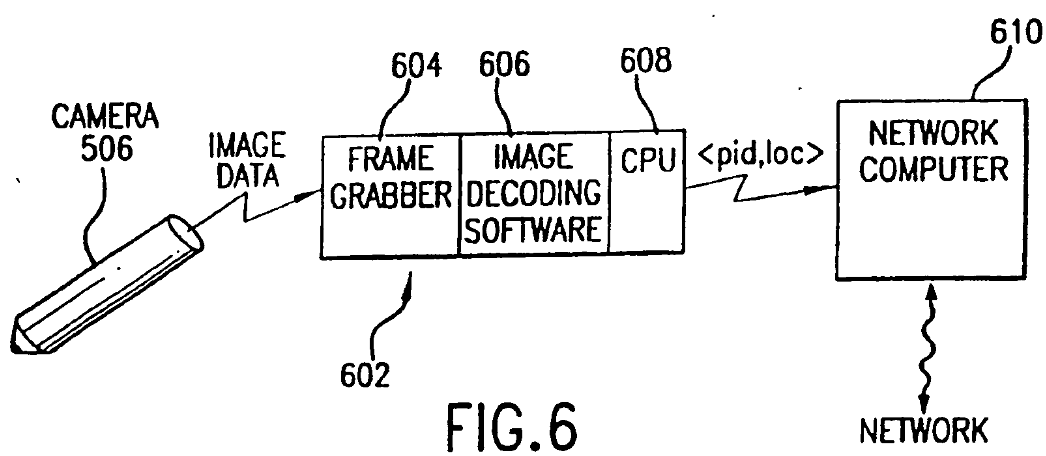

CAMERA
506
IMAGE
DATA
604
606
608
610
FRAME
GRABBER
IMAGE
DECODING
SOFTWARE
CPU
<pid,loc>
NETWORK
COMPUTER
602
NETWORK

FIG.6

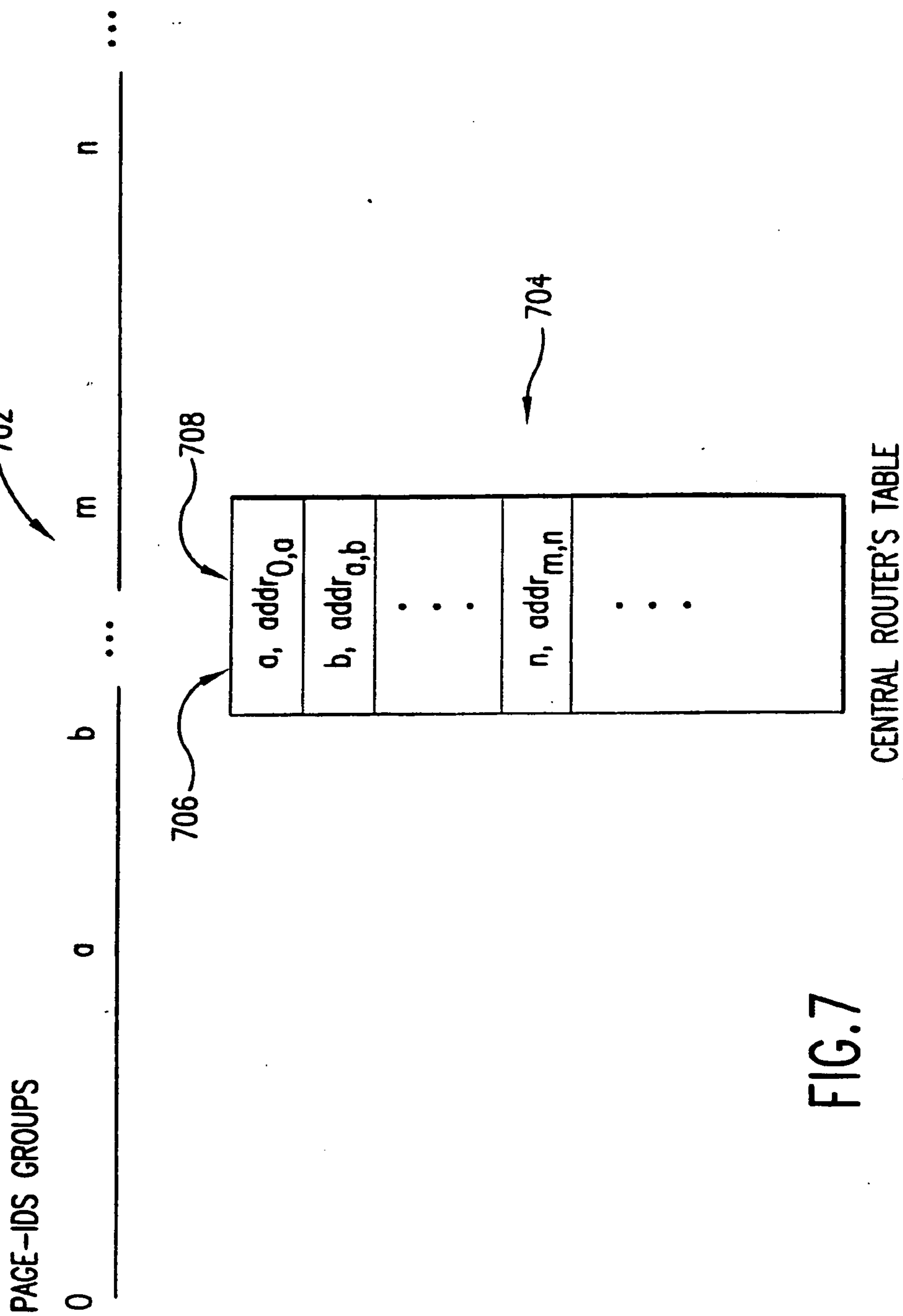

PAGE-IDS GROUPS
0
a
b
m
n
702
704
706
708
a, $addr_{0,a}$
b, $addr_{a,b}$
n, $addr_{m,n}$
CENTRAL ROUTER'S TABLE

FIG.7

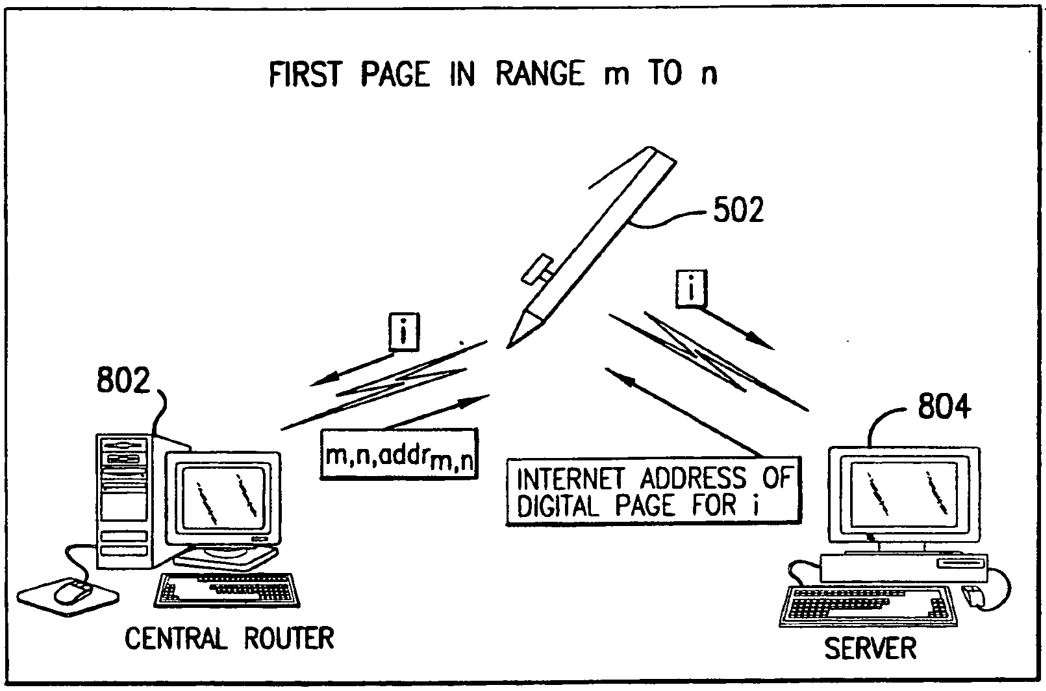

FIRST PAGE IN RANGE m TO n
502
i
i
802
804
m,n,addr$_{m,n}$
INTERNET ADDRESS OF DIGITAL PAGE FOR i
CENTRAL ROUTER
SERVER

FIG.8

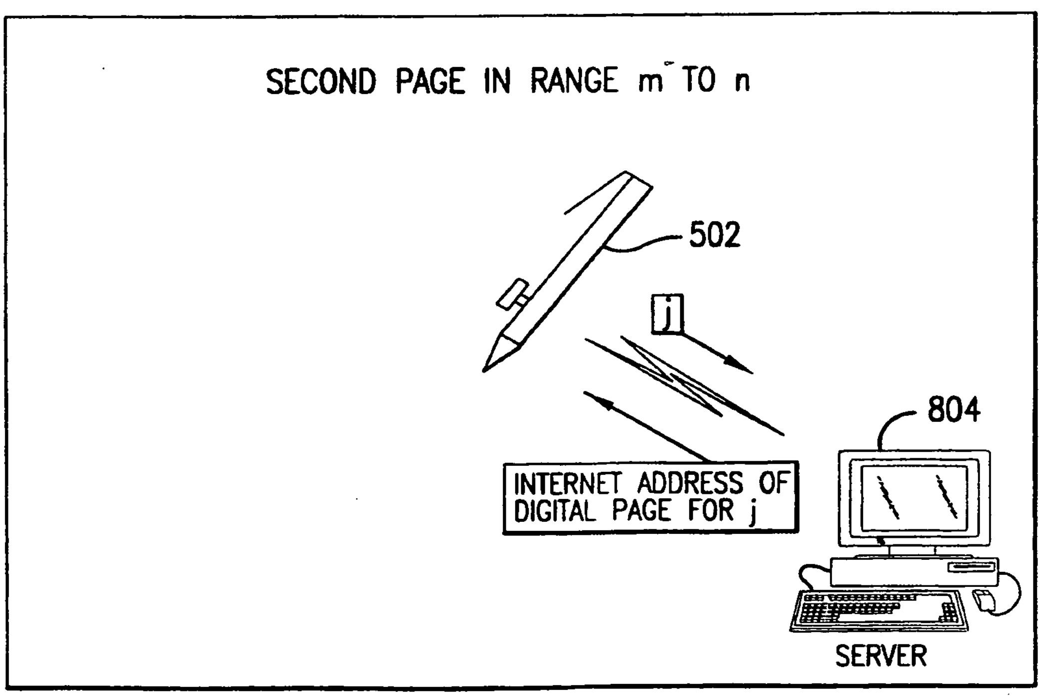

SECOND PAGE IN RANGE m TO n
502
j
804
INTERNET ADDRESS OF DIGITAL PAGE FOR j
SERVER

FIG.9

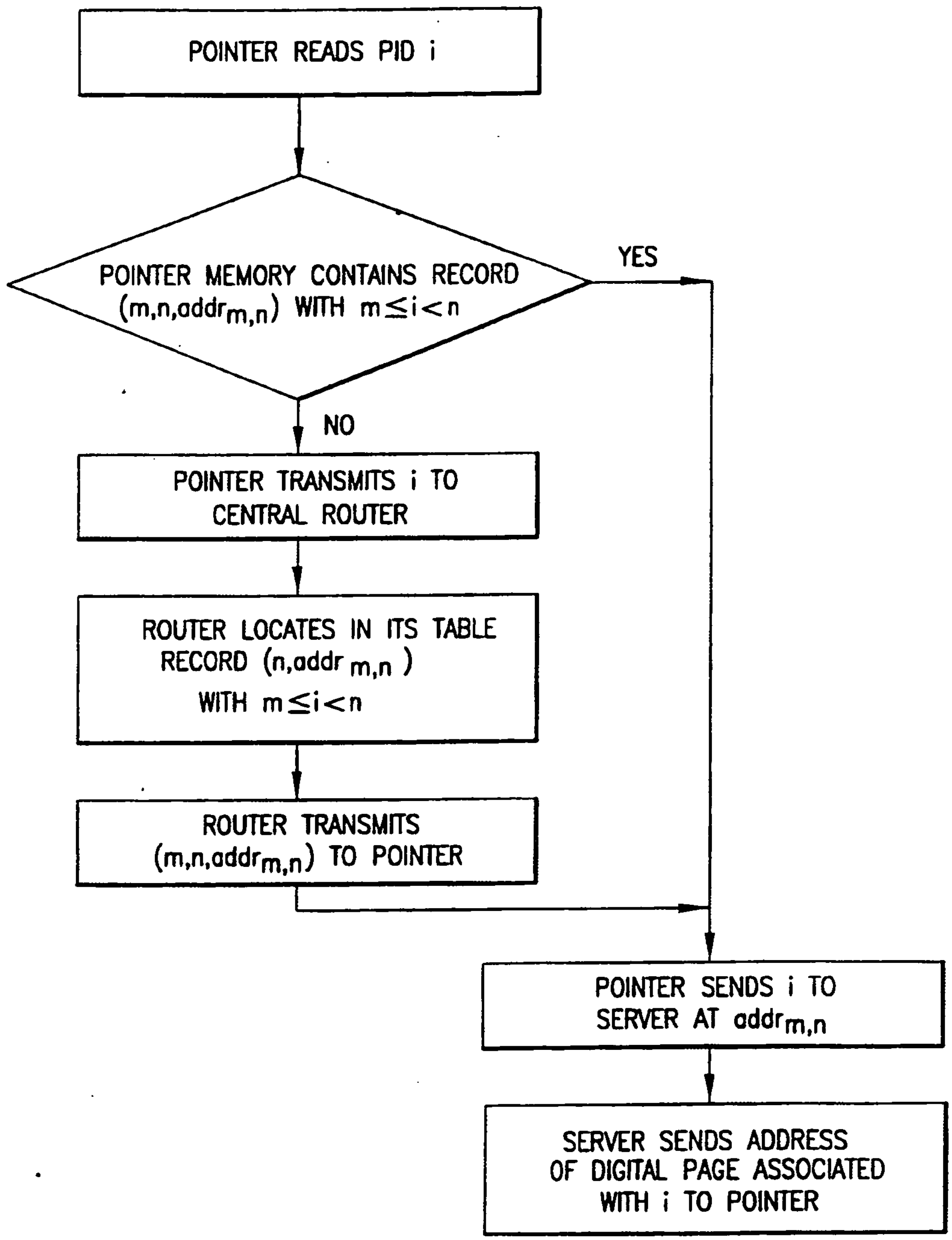

POINTER READS PID i
POINTER MEMORY CONTAINS RECORD
(m,n,addr$_{m,n}$) WITH m≤i<n
YES
NO
POINTER TRANSMITS i TO
CENTRAL ROUTER
ROUTER LOCATES IN ITS TABLE
RECORD (n,addr$_{m,n}$)
WITH m≤i<n
ROUTER TRANSMITS
(m,n,addr$_{m,n}$) TO POINTER
POINTER SENDS i TO
SERVER AT addr$_{m,n}$
SERVER SENDS ADDRESS
OF DIGITAL PAGE ASSOCIATED
WITH i TO POINTER

FIG.10

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/20596

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :G06F 12/00
US CL : 707/501
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 707/501; 79/88.08; 370/236

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 5,729,594 A (KLINGMAN) 17 March 1998, see abstract	1
Y	US 5,727,159 A (KIKINIS) 10 March 1998, see abstract	1
Y	US, 5,708,845 A (WISTENDAHL et al.) 13 January 1998, see abstract	1
Y	US 5,663,748 A (HUFFMAN et al.) 02 September 1997, see abstract	1

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
07 JANUARY 1999	04 MAY 1999

Name and mailing address of the ISA/US	Authorized officer
Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	ANTON W. FETTING
Facsimile No. (703) 305-3230	Telephone No. (703) 305-8449

Form PCT/ISA/210 (second sheet)(July 1992) ★